68-01-5971

A STUDY OF SAMPLING PROCEDURES AS APPLIED TO THE MECHANICAL INTEGRITY TESTING OF INJECTION WELLS

Submitted to

Dr. Jentai Yang
Office of Drinking Water

Mr. Thomas F. Sullivan
Contract Operations

Prepared for the
U.S. Environmental Protection Agency

by

Booz, Allen and Hamilton Inc.

Under the Direction of

Geraghty & Miller, Inc.

April 30, 1980

ACKNOWLEDGEMENT

This report was prepared for the Office of Drinking Water by Steve Heffernan of Booz, Allen and Hamilton with support from Walter Mardis, Walter Holman, and Jeff Mahan. Geraghty & Miller also assisted on questions related to injection well technologies. The EPA Task Manager was Arnold Kuzmack

TABLE OF CONTENTS

EXECUTIVE SUMMARY

1. AN INITIAL MECHANICAL INTEGRITY TEST (MIT) "CENSUS" OF ALL WELLS IS FAVORED OVER TESTING A SAMPLE OF WELLS

The use of sampling has been carefully considered from both a statistical analysis viewpoint and a technical viewpoint. After weighing the cost savings versus uncertainty trade-offs, sampling of injection wells is not considered a viable alternative to the proposed regulations at this time. This recommendation is consistent with EPA requirements. Some form of sampling procedure may be considered as a reasonable procedure later on for certain types of wells. Initial sampling is not considered a wise course of action for several reasons:

- A sampling design may miss some failed wells. A single failed well can jeopardize the safety of drinking water drawn from an aquifer in its vicinity. Since a single failed well can pollute an aquifer, it is undesireable to allow failing wells to go unnoticed. Although sampling may yield accurate indications (or estimates) of what population parameters are, it is not a solution to finding all failed wells.

- There are no concrete data on prior failure rates or prime causes of well failure. Such data is vital to the design of a sampling methodology. Although some causal data and failure rate data exist, the data are not comprehensive for all types of injection wells.

- Each class of well is unique. Class I wells have very different attributes from certain Class II and Class III wells. It is difficult to conceptualize sampling across a population of wells of such heterogeneity. Secondly, within each Class of well there can exist marked differences in age of well, depth, design, etc. Such differences greatly complicate the sampling task.

  - It is recommended that all Class I wells undergo initial MIT. [illegible] small (about 400), and (2) the injecta typically associated with Class I wells is often hazardous.

- It is further recommended that all Class II wells, with the exception of gas storage wells, be required to undergo initial MIT. Although the large number of Class II wells seems to encourage the use of sampling, excessive well heterogeneity - even within a given field - presents a significant barrier to sampling. EPA may wish to exempt gas storage wells from census testing as they are continually monitored.

2. <u>MID-COURSE EVALUATION DATA SHOULD BE GATHERED IN A FASHION APPROPRIATE FOR WELL-FAILURE ANALYSIS</u>

Mid-course evaluation data should form the backbone of any policy recommendations regarding the sampling of wells or the testing of all wells of a given type.

3. <u>STATISTICAL ANALYSIS OF MID-COURSE EVALUATION DATA MAY LEAD TO CHANGES IN WELL-TESTING POLICY</u>

Analysis of mid-course data may suggest changes in the timetable for well MIT. Exempting certain wells from testing or reducing the testing cycle is best accomplished through analysis of quantitative evaluation data, rather than qualitative "informed judgement."

4. <u>SAMPLE SIZE VARIES BY ASSUMPTIONS USED AND TYPE OF SAMPLING METHODOLOGY</u>

A sample size based on simple random sampling may require testing of only about 750 wells. Varying certain assumptions related to simple random sampling can increase the requirement to around 1500 wells. Stratified sampling, stratified by class of well, requires a sample size of about 5450 wells (See Table 7, Appendix).

The above sampling numbers represent [illegible] size required to ensure that the sample statistic (failure rate) is statistically close to the population parameter (failure rate). Note that sampling enables us to predict the failure rate within a defined margin of error, without having to test every well. It [illegible], on the other hand, [illegible] every failed well.

<u>A STUDY OF SAMPLING PROCEDURES AS APPLIED TO THE MECHANICAL INTEGRITY TESTING OF INJECTION WELLS</u>

The U.S. Environmental Protection Agency (EPA) has proposed regulations governing the mechanical integrity testing of injection wells.* Class I, II, and III injection wells will be required to undergo mechanical integrity testing (MIT) every five years, at a minimum. States with more stringent testing intervals will retain their stricter standards. The EPA regulations will stand as a default value for states where no MIT is currently required. Because MIT is a nontrivial expense the well owner incurs for every active injection well, EPA has considered the use of sampling to lessen the economic burden. Sampling introduces uncertainty** in terms of impact to fresh water aquifers. Since "uncertainty" is difficult to quantify, it is difficult to assess what an acceptable level of uncertainty is.

This report addresses the primary question, "Is it safe to allow a sample of injection wells to be tested in lieu of testing every one, as now proposed?" A secondary question, also addressed herein, is, "Is sampling a viable alternative later in the MIT process?" That is, if initially it is unwise to allow wells to go untested, would it be advisable to do so later?

---

* 40 CFR, Parts 122, 123, 124, and 146

** The term <u>uncertainty</u> is used in lieu of <u>risk</u>. Risk exists when the probability distribution for all possible outcomes is known. Uncertainty is a condition in which the probability distribution of all possible outcomes is <u>not</u> known. As is pointed out later in this report, the sampling of injection wells is a condition of uncertainty since we do not know [illegible] the probability distribution of failure.

1. TESTING A SAMPLE OF INJECTION WELLS IN LIEU OF TESTING ALL INJECTION WELLS IS NOT A VIABLE POLICY OPTION AT THIS TIME

There are several reasons why an initial sampling effort does not seem warranted. They are enumerated below.

(1) A Sampling Design May Miss Some Failed Wells

The object of MIT is to detect failed wells and flag them for rehabilitation or repair. A failed well can potentially pollute a potable water aquifer from which drinking water is drawn. A single failed well can jeopardize the safety of drinking water drawn from an aquifer in its vicinity. Sampling may yield a fairly accurate indication of the failure rate for a given type of well. It does not, however, help locate all failed wells so they may be repaired or replaced.

(2) There Are No Concrete Data on Prior Failure Rates Or Prime Causes of Well Failure

The above data are required if any form of sampling is to be utilized. Prior failure rate estimates range from a low of one percent* to 3.75 percent** for specific types of wells. These are only estimates as no formal system for tabulating this useful statistic exists at present.

Sampling recommendations are based on prior failure rate, error variance, well population size, and confidence level. For injection wells, prior failure rate is either unknown or uncertain, and we are not absolutely certain about the popululation N. Without such information, it is difficult to select a sampling frame or have much confidence in its suitability.

* Comments from [illegible], California [illegible], Dec., 197[illegible].

** Failure rate for [illegible] well [illegible] packer, from [illegible] of Compliance - [illegible] Control Program, [illegible].

The Appendix of this document contains computed sample size data for each class of well. Note that appropriate sample size varies relative to the values we assign when computing it. The various sensitivities are displayed for comparision.

Types of well failure are not adequately documented. Historical information regarding the specific corrective action taken on each failed well probably exists for some wells but has not been centrally organized, compiled or analysed. Knowning the types of well failure is an important first step in determining the underlying causes of well failure. In some instances, the reason for a well failure is obvious, such as a clear separation of the packer. However, whether the packer separation occurred because of excessive injection pressures, the age of the well, or some other variable, may not be as obvious.

(3) Each Class of Wells is Unique. Within Each Class There can be Great Variations

One of the reasons simple random sampling is not an available option in the testing of wells has to do with the heterogeneity of wells. There are three relevant classes with respect to injection wells. Each class is likely to have great differences even within the same field of wells. This situation has occurred when various sections of a given well field were drilled at different points in time. The oldest injectors may have been quite shallow, newer wells much deeper. If depth were a key variable in explaining the variance in failure rates, the same types of well in the same field might have very different likelihoods of failure, other things constant. The central theme is: similarly classified wells may have very different rates of failure, hence sampling a few may not give a true indication.

Each of the three classes of wells is "profiled" below according to their adaptability to sampling either now or later. We will consider the profile data later when considering alternatives to sampling.

- Class I Wells - Class I wells generally are used for disposal of industrial and municipal wastes in saline aquifers. Because of the toxic nature of the injectate, Class I wells are typically the [illegible] [illegible]. [illegible]

1 well per site, and a permit for every well. Few well failures in saline aquifers have been observed due to strict regulation and permit systems in states permitting Class I wells.

At least 404 industrial and municipal wastewater injection wells have been constructed in 25 states, at least 209 of which are operational. Nearly 60% are used by the chemical, petrochemical and pharmaceutical industries. Industrial injection rates are relatively low. Most inject less than 100 gpm (6 litre/sec). Municipal rates are higher (5-10 million gallons/day). Receiving reservoirs are distributed between sand, sandstone, and carbonate rocks; the three most common aquifer types. Because of the toxic chemical concentrations often present in industrial wastes, injection zones are usually deep. Only six percent are less than 1000 feet in depth. The majority are between 2000 and 6000 feet.*

For Class I injection wells, no type of sampling or exemption is felt warranted at this time. Because of the toxicity of injecta and low number of such wells, it is felt most appropriate to require an initial MIT of all Class I wells. Over time, some form of exemption criteria may emerge to lessen the number of Class I wells that need to be tested.

- <u>Class II Wells</u> - Class II injection wells are used for oil and gas storage and oil and gas production. Oil and gas production wells include enhanced recovery wells and brine disposal wells. Oil and gas storage wells vary from 1000 to 3000 feet in depth, their most common depth being 2000 feet. Wells associated with oil and gas recovery can vary from 1000 to 15,000 feet in depth, but are usually [illegible] 5000 feet deep.** Many Class II wells are converted producer wells which

* [illegible]

** [illegible]

have exhausted the oil field in which they are situated. While the majority are converted wells, the proportion varies from 90% converted wells in the Illinois Basin and Appalacia to a low of 60% converted wells in the Gulf Coast.* Table 1 below lists the Class II injection well population (1979) by region.

- <u>Class III Wells</u> - Class III wells are those used to inject fluids for the solution mining of minerals, for in-situ gasification and liquefaction of oil shale and coal, to recover geothermal energy, and wells for Frasch process sulfur mining. Well depths vary not only by type of Class III well, but among wells of the same type: Frasch sulfur wells range from 300 to 2000 feet in depth, salt solution mining from 200 to 10,000 feet, geothermal wells from 100 to 3,000 feet, oil shale from 300 to 1,200 feet. With the exception of uranium solution and copper mining, the toxicity of <u>injected</u> fluids is relatively low. The nature of fluid varies by application. The toxicity of produced fluids is moderate to high, however.**

Table 2, below, shows the number of Class III wells, although precise numbers of these wells by state are not available at this time. Certain short-lived Class III wells are exempt from the proposed five-year testing interval. All new and existing salt and geothermal wells will be required to undergo initial testing and subsequent testing at five-year intervals. Unlike Class II wells which may have been operational for decades, Class III wells may last a few weeks to 15 years. Uranium wells are usually only active one to two years. Copper solution mining, oil shale, coal, lignite, and tar sands injection wells last between two and three years. Salt solution mining wells may last ten to fifteen years and geothermal sites may be productive for fifteen to thirty years.

* [illegible]

** [illegible]

TABLE 1

CLASS II INJECTION WELL POPULATION DATA BY GEOGRAPHIC REGION

Regions	Salt Water Disposal: Wells	Salt Water Disposal: % of Total	Enhanced Recovery: Wells	Enhanced Recovery: % of Total
	Projected Number of Existing Injection Wells as of December 31, 1979			
1. Illinois Basin	6,855	17.4%	12,387	12.3%
2. Appalachia	5,789	14.7	5,752	5.7
3. Mid-Continent	5,365	13.6	30,027	29.9
4. Permian Basin	5,726	14.5	26,600	26.5
5. Gulf Coast	6,921	17.6	1,104	1.1
6. East Texas	5,273	13.4	1,840	1.8
7. Rocky Mountain	158	0.4	3,517	3.5
8. California	545	1.4	14,861	14.8
Total Wells in Region Studied	36,632	93.0	96,088	95.6
Total Wells in Other Region	2,723	7.0	4,227	4.4
Total Wells in U.S.A	39,355	100.0%	100,315	100.0%

Source: Arthur D. Little, Estimates

TABLE 2

ESTIMATED NUMBER OF CLASS III
SPECIAL PROCESS INJECTION WELLS

Type of Well	Sites (fields)	Number 1979/80)	Projected 1985	Location
1. Sulfur mining (Frasch process)	8-10	500	500-600	TX, LO
2. Solution mining				
a. Uranium	33	6,300	18,000	WY, TX, NM, CO
b. Salt	80	1,000	1,100	NY, WV, PA, TX, LO, KA
c. Copper & other metals		10-20	30-50	CO, UT, MI, AZ
3. In Situ Gasification & Liquefaction	7	30	300	CO, UT, WY, TX, SD ND, MT, CA, OR NM, ID
4. Geothermal	6	25	50	CA
		7700	20,000	

Geraghty & Miller, "Development of Procedures for Sub-classification of Class III Injection Wells", January 7, 1980, Draft Final Report.

Of all Class III wells, Frasch sulfur and salt solution wells seem best suited for sampling. Within a given state, wells of the above variety are predominantly homogeneous. The Frasch sulfur process calls for many wells, similar in design, to be dug in a new field. A field is then mined as rapidly and completely as possible. Once the field is depleted, the wells are pulled up, and a new field is exploited. Only about one-third of the well casing comes up as the self-sealing nature of the process "cements" the bottom in the well. Because wells within a field are virtually of the same design, depth, and age, a sample of such wells is likely to yield statistics very close to true population parameters. Hence, sampling incurs less uncertainty for these types of wells than other Class III wells. Further analysis is needed to determine if other Class III wells are as well-suited for sampling.

2. MID-COURSE EVALUATION DATA SHOULD BE GATHERED AND RETAINED FOR POTENTIAL STATISTICAL ANALYSES IN THE FUTURE

Mid-course data, gathered nationwide, could be useful for certain analyses of well data. If EPA deems such analyses appropriate, mid-course data should be gathered in a fashion which makes possible statistical analysis of collected data. These data may indicate causal factors in well failure and form the basis for changes in well testing policy. The methods of collecting data, or its usefulness, must ultimately be decided by the EPA.

APPENDIX

This appendix contains sample size data for all classes of wells considered in this analysis. While several strong objections to initial sampling have been raised, there remains considerable interest in sampling statistics, should sampling become a viable alternative in the future. Accordingly, well population information has been evaluated and estimates of sample size drawn from that information. Each of the primary input criteria is varied, holding other items constant to show the various sensitivities.

The broadest possible sampling scenario would treat all injection wells as having the same rate of failure (expressed in this context as probability of failure), and would involve a simple random sample drawn from the entire well population. This approach has the following advantages:

- Lowers front-end costs of MIT
- Lowers time required to perform total MIT.

Its disadvantages are as follows:

- Ignores gross differences in well types
- Relies on a single estimate of failure rate
- May leave polluted aquifers undiagnosed
- Does not allow for comprehensive data collection
- State-of-the-art remains one of uncertainty, as opposed to risk.

Sample size is derived as follows. Assuming the population is normally distributed and wells are randomly sampled, we can generate an estimate of total sample size (n) as follows:

$$n = \frac{N Z^2 p^*(1-p^*)}{(N-1)(e^2) + Z^2 p^*(1-p^*)} \quad (1)$$

where: N = population size estimate

[illegible]

p* = prior probability of an event occurring (event = well failure) or best-guess if unknown

e = error term (margin of allowable error in predicting p*)

The first data item, population size, must be estimated. Figures drawn from Tables 1, 2, and 3 give us the following estimate for N:

Class I:	209
Class II:	
SWD	39,355
ER	100,315
Class III:	
Frasch	500
Solution Mining	7,320
Gas & Liquid	30
Geothermal	25
	N = ~150,000

The EPA "Guide to the UIC Program" reports, "It is estimated that perhaps as many as 500,000 injection wells are in operation nationwide."* Both numbers are used in the analysis for comparitive purposes. The default value will be 150,000 wells.

The second datum is the Z statistic or alpha level. Several alpha levels are considered and their effect on sample size is noted. Z values in this exercise vary from 1.282 to 2.576. The default value will be 2.576.

The third item of information, p*, is a "guesstimate" value of the proportion of wells which fail—the failure rate. We do not know what value p* takes on. Estimates range from a low of 1 for certain oil and gas related injections to a high of 3.75 for certain enhanced recovery wells. A range of failure rate estimates from one to four percent is used, with two percent as a default value.

* [illegible]

The last item needed is the error term. It represents what the analyst considers an acceptable margin of error in predicting p, the failure rate. The error term is inversely proportional to sample size. Various error terms will be tried, relying on a default value of one-half of p*, or one percent.

1. <u>SAMPLE SIZE UNDER THE SIMPLE RANDOM SCENARIO CAN VARY BETWEEN ABOUT 500 to 1,500</u>

TABLE 4

Sample Size for Varying
Population N Assumptions,
$\alpha = .05$, $p^* = 2\%$, $e = 1\%$

POP SIZE	SAMPLE SIZE
150,000	749
300,000	751
600,000	752

Table 4 indicates that rather large changes in the overall well population size produce relatively small changes in sample size, other things constant. Table 5 below, shows the effect on sample size when the confidence level, α is varied. The α level of .05, for example, should be interpreted as, "95 out of 100 times we expect the sample statistic to fall within the probability distribution for the population parameter." A smaller α level improves sampling precision.

TABLE 5

Sample Size for Varying
Alpha levels, N = 150,000
$p^* = 2$[illegible], $e = 1$[illegible]

LEVEL	SAMPLE SIZE
.10	[illegible]
.05	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]

Table 6 shows the range of sample sizes for various estimates of well failure rate, assuming one rate is chosen to represent all wells. Note that in Table 6, <u>two</u> values are being allowed to change: the failure rate <u>and</u> error term. The error term is defined by the failure rate in each instance. Column n' shows the required sample size if the error term is kept at a constant 1%.

TABLE 6

Sample Size for Varying
Prior Failure Rate Estimates
$N = 150,000$, $\alpha = .05$, $e = p^*/2$

FAILURE RATE	SAMPLE SIZE	n'
1%	1,506	379
2%	749	749
3%	495	1110
4%	368	1461

2. <u>STRATIFYING BY CLASS OF WELL AND RANDOM SAMPLING EACH CLASS PRODUCES THE FOLLOWING SAMPLE SIZES:</u>

TABLE 7

Estimates of Sample Size, Stratified by
Class of Well for Known Well Populations,
$\alpha = .01$, $e = .01$, $p^* = 2\%$

	Pop N	Sample N
Class I	209	[illegible]
Class II		
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
Class III		
Frasch	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]

Note that stratifying by class and type of well and drawing a random sample from each group increases the overall sample size to approximately 5450, instead of the 1250 for a simple random sampling. Stratified sampling has the following advantages over random sampling:

- Acknowledges differences in types and Classes of wells
- Is likely to find more failed wells.

Neither stratified or random sampling is advised unless a census of all wells has first occurred. Once a census of wells has taken place, some form of sampling is attractive because:

- It lowers MIT program life-cycle costs
- It allows MIT to be more easily administered.